PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-077177
(43)Date of publication of application : 22.03.1996

(51)Int.Cl. G06F 17/30

(21)Application number : 06-208308
(22)Date of filing : 01.09.1994

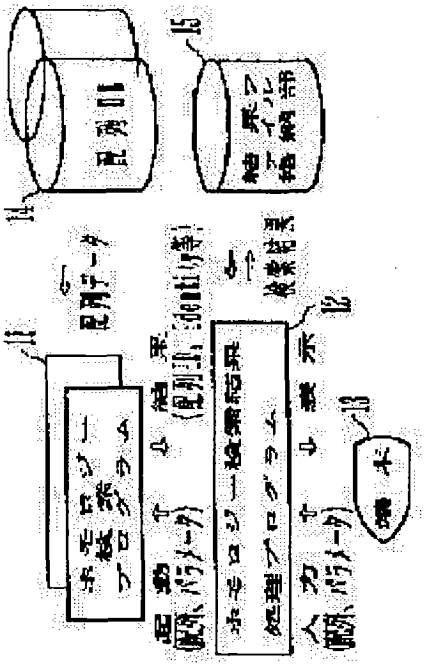
(71)Applicant : FUJITSU LTD
(72)Inventor : NOGUCHI TAMOTSU

(54) LIST PROCESSING SYSTEM AND METHOD THEREFOR

(57)Abstract:

PURPOSE: To provide a system and a method for efficiently processing plural lists with plural data, and extracting a feature such as a point of similarity or a point of difference, etc., among those.

CONSTITUTION: A homology retrieving program 11 retrieves an array data base 14 in which the known array data of a gene or protein is stored; and outputs plural lists in which the identifiers of the array data similar to the array data of a retrieved object are arranged in the order of the highness of similarity. A homology retrieved result processing program 12 adds information used in the preparation of these lists to the list as a file name, and stores it in a result file storage part 15 together with the list. Then, the homology retrieved result processing program 12 compares the contents of plural lists, and extracts the feature such as the point of similarity or the point of difference among those, and displays it on the display of a terminal 13.



LEGAL STATUS

[Date of request for examination] 08.08.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（A）**　(11)特許出願公開番号

特開平8－77177

(43)公開日　平成８年(1996)３月22日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　17/30				
		9194－5L	Ｇ０６Ｆ　15/ 401　３２０　Ｚ	

審査請求　未請求　請求項の数22　ＯＬ　（全 24 頁）

(21)出願番号　特願平6－208308

(22)出願日　平成６年(1994)９月１日

(71)出願人　000005223
富士通株式会社
神奈川県川崎市中原区上小田中1015番地
(72)発明者　野口　保
神奈川県川崎市中原区上小田中1015番地
富士通株式会社内
(74)代理人　弁理士　大菅　義之　（外１名）

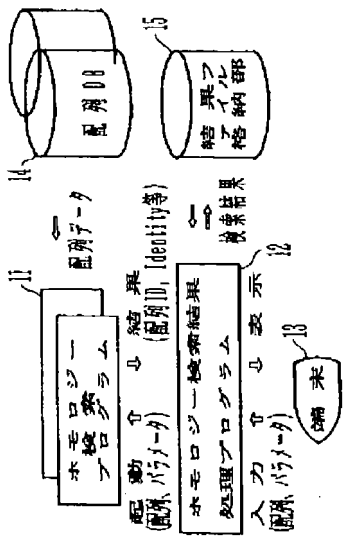
(54)【発明の名称】　リスト処理システムとその方法

(57)【要約】

【目的】　複数のデータを有する複数のリストを効率的に処理して、それらの間の類似点や相違点等の特徴を抽出するシステムとその方法を提供する。

【構成】　ホモロジー検索プログラム１１は、遺伝子や蛋白質の既知の配列データを格納している配列データベース１４を検索して、検索対象の配列データに類似した配列データの識別子を類似度の高い順に並べたリストを複数出力する。ホモロジー検索結果処理プログラム１２は、これらのリストの作成に用いられた情報をファイル名としてリストに付加し、リストとともに結果ファイル格納部１５に格納する。そして、ホモロジー検索結果処理プログラム１２は複数のリストの内容を比較し、それらの間の類似点や相違点等の特徴を抽出して端末１３のディスプレイ上に表示する。

実施例の構成図



1

【特許請求の範囲】

【請求項1】　データベースに格納されたデータを処理する情報処理装置において、
前記データベースの検索結果であって複数のデータより構成されるリストを複数個格納するリスト格納手段と、
前記リスト格納手段に格納された前記複数のリストについての特徴を抽出して出力する特徴抽出手段と、
を備えたことを特徴とするリスト処理システム。

【請求項2】　前記リスト格納手段は、前記複数のデータに順位を付加した前記リストを格納することを特徴とする請求項1記載のリスト処理システム。

【請求項3】　前記リスト格納手段は、前記データに順位を付加するときに用いた情報を前記リストのファイル名に付加して格納することを特徴とする請求項2記載のリスト処理システム。

【請求項4】　前記特徴抽出手段が出力する前記特徴を画面表示する特徴表示手段をさらに備えたことを特徴とする請求項1記載のリスト処理システム。

【請求項5】　データベースに格納されたデータを検索して、与えられたデータに類似したデータの識別子を出力するホモロジー検索装置において、
複数の前記類似したデータの識別子を有するリストを複数個格納するリスト格納手段と、
前記リスト格納手段に格納された前記複数のリストについての特徴を抽出して出力する特徴抽出手段と、
を備えたことを特徴とするリスト処理システム。

【請求項6】　前記リスト格納手段は、前記複数の類似したデータの識別子に順位を付加した前記リストを格納することを特徴とする請求項5記載のリスト処理システム。

【請求項7】　前記リスト格納手段は、前記類似したデータの識別子に順位を付加するときに用いた情報を前記リストのファイル名として格納することを特徴とする請求項6記載のリスト処理システム。

【請求項8】　前記リスト格納手段は、前記与えられたデータの識別名と、前記類似したデータの識別子に順位を付加するときに用いた手法の識別名と、前記データベースの識別名と、前記手法のパラメータのうちのいずれかを前記リストのファイル名に付加して格納し、前記複数のリストを管理することを特徴とする請求項7記載のリスト処理システム。

【請求項9】　前記特徴抽出手段は、前記複数のリストの間の類似点と相違点のうちいずれかを前記特徴として抽出することを特徴とする請求項5記載のリスト処理システム。

【請求項10】　前記特徴抽出手段は、前記複数のリストに共通に含まれる識別子を前記特徴として抽出することを特徴とする請求項5記載のリスト処理システム。

【請求項11】　前記特徴抽出手段は、前記複数のリストに共通に含まれない識別子を前記特徴として抽出することを特徴とする請求項5記載のリスト処理システム。

2

【請求項12】　前記特徴抽出手段は、前記複数のリストのうち指定されたリストに含まれない識別子を前記特徴として抽出することを特徴とする請求項5記載のリスト処理システム。

【請求項13】　前記特徴抽出手段は、前記複数のリストにおいて指定された識別子を前記特徴として抽出することを特徴とする請求項5記載のリスト処理システム。

【請求項14】　前記特徴抽出手段は、前記複数のリストに共通に含まれる識別子のうち同順位の識別子を前記特徴として抽出することを特徴とする請求項5記載のリスト処理システム。

【請求項15】　前記特徴抽出手段は、前記複数のリストに含まれる識別子の数を前記特徴として抽出することを特徴とする請求項5記載のリスト処理システム。

【請求項16】　前記特徴抽出手段が出力する前記特徴を画面表示する特徴表示手段をさらに備えたことを特徴とする請求項5記載のリスト処理システム。

【請求項17】　前記リスト格納手段は、前記類似したデータの識別子に順位を付加するときに用いた情報を前記リストのファイル名として格納し、
前記特徴表示手段は、前記複数のリストの前記ファイル名を画面表示し、さらに指定されたファイル名を持つリストの内容を画面表示することを特徴とする請求項16記載のリスト処理システム。

【請求項18】　前記特徴表示手段は、前記特徴をグラフ表示することを特徴とする請求項16記載のリスト処理システム。

【請求項19】　データベースに格納されたデータを処理する方法であって、
前記データベースの検索結果であって複数のデータより構成されるリストを複数個格納し、
前記データベースを検索するときに用いた情報を前記リストのファイル名に付加して格納し、
前記ファイル名を用いて前記複数のリストを前記情報の項目別に分類して管理することを特徴とするリスト処理方法。

【請求項20】　前記リストを構成する前記複数のデータに順位を付加し、
前記複数のデータに順位を付加するときに用いた情報を前記リストのファイル名に付加して格納することを特徴とする請求項19記載のリスト処理方法。

【請求項21】　格納された前記複数のリストの前記ファイル名を表示し、
表示された前記ファイル名のうちの複数のファイル名を前記情報を用いて選択し、
選択された前記複数のファイル名に対応する複数のリストの内容を表示することを特徴とする請求項19記載のリスト処理方法。

【請求項22】　選択された前記複数のファイル名に対

3

応する複数のリストについての特徴を抽出し、
抽出された前記特徴を画面表示することを特徴とする請求項２１記載のリスト処理方法。
【発明の詳細な説明】
【０００１】
【産業上の利用分野】本発明は、複数のデータを有する複数のリストを処理して、それらの間の特徴を抽出するリスト処理システムとその方法に関する。
【０００２】
【従来の技術】近年のバイオテクノロジーの進歩が社会に及ぼす影響は、日増しに増大しつつある。特に、遺伝子の本体であるデオキシリボ核酸（ＤＮＡ）分子に組み換え等の操作を施す遺伝子工学や、既存の蛋白質を基にして新しい蛋白質を作り出す蛋白質工学の発達には著しいものがある。
【０００３】ＤＮＡは、塩基、糖（デオキシリボース）、リン酸から成るヌクレオチドと呼ばれる構成単位から成る高分子である。このうち、ＤＮＡを構成する塩基には、アデニン（Ａ）、チミン（Ｔ）、シトシン（Ｃ）、グアニン（Ｇ）の４種類がある。それぞれのヌクレオチド間では、デオキシリボースとリン酸同士が結合して鎖状に繋がり、２本の鎖が互いに塩基で結びついた二重螺旋構造を持つ。
【０００４】塩基同士の結びつきには規則性があり、ＡとＴ、ＣとＧがそれぞれ結合する。これらの塩基のＤＮＡ内における配列順序がＤＮＡの種類、すなわち遺伝子の種類を決めている。
【０００５】このようにＤＮＡの塩基配列が遺伝情報を記述しているため、遺伝子工学においては、与えられたＤＮＡの複雑な塩基配列（遺伝子配列）を正確かつ迅速に解析する技術が必要不可欠である。
【０００６】蛋白質は、種類の異なる多数のアミノ酸がペプチド結合により鎖状に連結した高分子化合物である。アミノ酸だけからできているポリペプチドは単純蛋白質と呼ばれ、アミノ酸と核酸、炭水化物、リン酸等が結合したものは複合蛋白質と呼ばれる。蛋白質の多様な機能は、ポリペプチド鎖を作るアミノ酸の配列順序やポリペプチド鎖の幾何学的配置等により決められている。したがって、蛋白質工学においては、与えられた蛋白質のアミノ酸配列を正確かつ迅速に解析する技術が必要となる。
【０００７】従来より遺伝子や蛋白質の性質を調べるために、その塩基配列やアミノ酸配列をデータベースに格納された既知の配列データと比較して、与えられた遺伝子等との間にホモロジー（相同性）を有する配列データを求める手法が採られている。このようにデータベース内の類似配列を検索する手法はホモロジー検索と呼ばれている。ホモロジー検索において、一般的には比較する２つの配列データの先頭からそれらの類似部分を検索していく。そして各部分の類似度を計算して配列全体としての類似性を評価する。

4

【０００８】しかし、このようなホモロジー検索の手法には確立されたものがあるわけではなく、様々な手法が併用され、比較されながら用いられている。また、検索のために用いられる既存のデータベースにも、複数の異なるものが存在する。同じ配列データを対象にしてホモロジー検索を行っても、その検索結果は用いた手法やデータベース、また、検索の際に与えるパラメータ等によって異なってくる。
【０００９】従来は、いくつかの手法やパラメータ等を組み合わせて何回かホモロジー検索を行い、それらの結果を比較して検索データの取捨選択を行ったり、最適な検索手法やパラメータを決定したりしていた。
【００１０】そして、個々のホモロジー検索結果をリスト形式で出力し、手作業でそれらの結果の間の類似点や相違点を調べていた。
【００１１】
【発明が解決しようとする課題】しかしながら従来の作業方法においては、ホモロジー検索結果が少ない場合は手作業で対処できるが、その数が増えるにしたがって時間もかかるし、間違いも多くなるという問題がある。
【００１２】また、最近の遺伝子及びアミノ酸配列自動読み取り装置の普及と、遺伝子情報の解明を目的としたヒトゲノムプロジェクトなどのビックプロジェクトの成果として、配列データの件数が飛躍的に増加したのに伴い、個々の検索結果に含まれるデータは既に人間の手作業では対処できない数に達している。
【００１３】本発明は、複数のデータを有する複数のリストを効率的に処理して、それらの間の特徴を抽出するリスト処理システムとその方法を提供することを目的とする。
【００１４】さらに詳しくは、バイオテクノロジー等の分野におけるホモロジー検索の結果を項目別に分類管理することにより、それらの結果の類似点や相違点等の特徴を明確に示すことを目的とする。
【００１５】
【課題を解決するための手段】本発明は、データベースに格納されたデータを処理する情報処理装置におけるリスト処理システムとその方法である。
【００１６】図１は本発明のリスト処理システムの原理図である。本発明のリスト処理システムは、リスト格納手段１、特徴抽出手段２、および特徴表示手段３を備え、データベース４に格納されたデータを用いて処理を行う。
【００１７】リスト格納手段１は、データベース４の検索結果であってデータベース４に格納された複数のデータより構成されるリストを複数格納し、データを検索するときに用いた情報を得られたリストのファイル名に付加して格納する。
【００１８】また、リスト格納手段１は、リストを構成

5

する複数のデータに順位を付加して格納し、データに順位を付加するときに用いた情報をそのリストのファイル名に付加して格納する。

【0019】データベース4がホモロジー検索のためのデータベースである場合は、リスト格納手段1は、検索により得られた、与えられたデータに類似したデータの識別子に順位を付加して得られるリストを複数格納する。またリスト格納手段1は、上記与えられたデータの識別名、上記類似したデータの識別子に順位を付加するときに用いた手法の識別名、データベース4の識別名、上記手法のパラメータ等の情報のうちいずれかを前記リストのファイル名に付加して格納し、複数のリストを管理する。

【0020】特徴抽出手段2は、リスト格納手段1に格納された複数のリストの間の類似点や相違点等の特徴を抽出して出力する。特徴表示手段3は、ウィンドウ機能およびグラフィック機能を有し、複数のリストのファイル名を並べて画面表示し、ユーザがいくつかのファイル名を指定すると、指定されたファイル名を持つリストの内容を画面表示する。また特徴表示手段3は、ウィンドウ機能またはグラフィック機能を用いて特徴抽出手段2が出力する上記特徴を画面表示する。

【0021】例えば図3に示されるように、本発明のリスト格納手段1は結果ファイル格納部15であり、特徴抽出手段2はホモロジー検索結果処理プログラム12を実行する不図示のプロセッサであり、特徴表示手段3はホモロジー検索結果処理プログラム12を実行するプロセッサと端末13である。

【0022】

【作用】リスト格納手段1が与えられたデータや手法の識別名等の情報をリストのファイル名に付加して管理し、特徴表示手段3がそれらの識別名等を付加されたファイル名を画面に表示するので、ユーザは各リストがどのようにして得られたのかを一目で認識できる。また、表示されたファイル名を持つリストの中から特定の情報を持つリストを容易に選ぶことができる。

【0023】例えば、ホモロジー検索により得られた配列データの配列名のリストに、検索対象の配列名や検索手法名等をファイル名として付加すれば、リスト（ファイル）の内容を見なくても、どのような検索結果が格納されたファイルであるかが分かる。

【0024】また、特徴抽出手段2が複数のリストの間の類似点や相違点等の特徴を抽出して、特徴表示手段3が抽出された特徴を分かりやすく画面に表示するので、複数のリストの間の特徴を効率よく把握することが出来る。ホモロジー検索結果の処理においては、従来のように検索された配列名を手作業で比較する必要がなく、検索結果に含まれる配列名が膨大な数になる場合や多数の検索結果のリストを比較する場合でも、高速にリストを処理することができる。グラフィック機能を用いて、抽

6

出された特徴をグラフ表示すれば、より明確に特徴を認識できる。

【0025】

【実施例】以下、本発明のバイオテクノロジーの分野における実施例について、図面を参照しながら詳細に説明する。

【0026】図2は、バイオテクノロジーの分野におけるホモロジー検索結果に影響を与える複数の項目を示している。図2において、配列は、検索の対象となる遺伝子の塩基配列または蛋白質のアミノ酸配列である。

【0027】ホモロジー検索手法としては、ＦＡＳＴＡ、ＢＬＡＳＴ（Basic Local Alignment Search Tool）、Smith-Waterman法等が知られている。ＦＡＳＴＡは、比較する2つの配列データを単位長さ毎に区切って、この比較単位をずらしながら類似度を計算する手法である。ＢＬＡＳＴはＦＡＳＴＡと似ているが、ＦＡＳＴＡより部分的な類似性を高く評価する手法であり、またＦＡＳＴＡよりも高速に結果が得られる。Smith-Waterman法は正確な結果が得られるが、複数のアルゴリズムを用いるため時間がかかる。

【0028】既知の配列データベースとしては、我が国のＤＤＢＪ（DNA Data Bank of Japan）、アメリカのGenBank（Genetic Sequence Data Bank）、ヨーロッパのＥＭＢＬ（European Molecular Biology Laboratory）のヌクレオチド配列データベース、ＮＢＲＦ（National Biomedical Research Foundation）の核酸配列データベース、SWISS-PROT（Swiss Protein Sequence Data Bank）等がある。

【0029】また、パラメータは、各ホモロジー検索手法において用いられるパラメータであり、その値の取り方によって同じ検索手法を用いても結果が変わってくる。当然の事ながら、これらの検索結果がユーザ毎に必要になり、それらを必要に応じて表示しなくてはならない。そこで結果ファイル群のデータをファイル名の指定により簡単に読み込むことができるようなシステムが必要になる。

【0030】図3は、本発明の実施例のホモロジー検索結果処理システムの構成を示している。図3のシステムは、既知の配列データを格納する配列データベース（配列ＤＢ）14、ホモロジー検索の結果得られた配列データのリストを格納する結果ファイル格納部15、入出力に用いられる端末13、プログラムを格納する不図示のプログラムメモリ、およびプログラムを実行する不図示のプロセッサを有する。ホモロジー検索プログラム11とホモロジー検索結果処理プログラム12は、プログラムメモリに格納され、プロセッサにより実行される。

【0031】ホモロジー検索はある配列に対して相同性を持った配列を配列データベース14中より探すもので、相同性の高い順に検索結果を出力する。その際の指

7

標となる数値は用いる手法によって異なるが、一般には同一性（Identity）と呼ばれる指標で表される。

【００３２】ホモロジー検索プログラム１１は、複数のホモロジー検索手法に対応する複数の検索プログラムから成り、ホモロジー検索結果処理プログラム１２から起動される。また、配列データベース１４は、複数の異なるデータベースを表している。

【００３３】検索の対象となる配列は、ユーザによる端末１３の操作により、キー入力、データベース検索、シーケンサ（配列読み取り装置）等の様々な方法で入力される。さらに端末１３よりホモロジー検索手法及びそのパラメータや検索を行うデータベースが指定されると、ホモロジー検索結果処理プログラム１２はその情報を指定されたホモロジー検索プログラム１１に渡し、指定された配列データベース１４内でのホモロジー検索を行わせ、その結果を受け取る。このとき受け取るホモロジー検索結果は、例えば検索された配列データの識別子（配列ＩＤ）やIdentity等である。

【００３４】ホモロジー検索結果処理プログラム１２は、ホモロジー検索結果を一旦結果ファイル格納部１５に格納し、次に指定された検索手法やパラメータに従ってホモロジー検索プログラム１１を起動し、同一の配列について検索を行わせる。これを繰り返すことにより、同一の配列を対象とした複数の結果ファイルが得られる。その後ホモロジー検索結果処理プログラム１２は、各結果ファイルの間の類似点や相違点等の特徴を抽出し、端末１３に表示する。抽出された結果ファイル間の特徴は、不図示のプリンタ等を用いて出力することもできる。

【００３５】本実施例のシステムはＵＮＩＸ環境を前提としており、各ファイルはＵＮＩＸの持つディレクトリ構造の中に作成される。結果ファイルは、図４に示すように、ユーザのホームディレクトリ配下の本システム専用のディレクトリｈｓの配下に作成される。

【００３６】結果ファイルのファイル名は、図４に示すように、図２の分類項目毎に区切られ、配列名、手法名、データベース（ＤＢ）名の各識別名、およびパラメータから成る。例えば図４のファイル名ＨＩＶ１１＿ＦＡＳＴＡ＿ＳＷ＿５．５．１のうち、ＨＩＶ１１は検索対象となる未知の配列名を表し、ＦＡＳＴＡは検索に用いる手法名を表し、ＳＷ（SWISS-PROT）は検索に用いるデータベース名を表し、５．５．１はＦＡＳＴＡで用いられるパラメータを表す。このような命名規約によりファイル名を管理すれば、どの配列についてどの手法、データベース、パラメータを用いて検索したのかがファイル名から一目瞭然となる。

【００３７】図５は、ディレクトリｈｓの配下に作成された結果ファイル名のリストの一例である。ＨＩＶ（Human Immunodeficiency Virus）、ＬＹＳＯ（Lysozyme）、ＵＢＩＱ（Ubiquitin）、ＬＥＣＴＩＮ、ＴＲＹ

8

ＰＳＩＮは検索対象の配列名を表し、ＦＡＳＴＡ１、ＦＡＳＴＡＮ、ＦＡＳＴＡＯは互いに異なる３種のＦＡＳＴＡを表し、ＳＭ－ＷＴはSmith-Waterman法を表す。ＳＷはデータベースSWISS-PROTを表し、５．２．１等はそれぞれの検索手法において用いたパラメータである。

【００３８】図５の各結果ファイルのデータ構造の一例が図６に示されている。図６の結果ファイルは、基本的にはそのファイル名に記述された情報の詳細な内容と、検索された配列に関する情報とから成っている。

【００３９】図６の例では、ファイル名に５．２．１と記されたパラメータが、ギャップペナルティ（Gap Penalty）Ｕ＝５、ギャップペナルティＶ＝２、ktup＝１であることを示しているのがわかる。ktupは、ＦＡＳＴＡにおいて一度に比較する部分配列に含まれる構成単位（塩基、アミノ酸等）の数を表す。例えば、ここでは構成単位を１個ずつ比較していくことを表している。ギャップペナルティＵ、ギャップペナルティＶについては後述する。

【００４０】また、検索対象（TARGET）の実際の配列名はＨＩＶ－１　PROTEASEであり、検索したデータベースはSWISS-PROTであることがわかる。ＬＩＳＴは、検索されたデータベース内の配列のエントリー名をそのスコアとともに、スコアの高い順に並べたリストである。配列のエントリー名は、図２の各データベース毎に決められた配列データの識別子あるいは識別名であり、図６ではSWISS-PROTのエントリー名として、例えばＨＩＶ　ＭＡＮＭＡ等が示されている。スコアは配列データのIdentityを表し、その値が大きいほど検索対象の配列に類似していることを示す。例えばＨＩＶ　MANMAのスコアは１１３３であり、この検索結果においてはＨＩＶ－１　PROTEASEに最も類似していると考えられる。

【００４１】図７は、ＦＡＳＴＡによるスコア計算に用いられるスコアテーブルの一例を示している。図７のスコアテーブルは、２つのアミノ酸の間の類似度を表すマトリクスであり、その行と列はそれぞれＡからＸまでのアミノ酸の名称を成分として持つ。Ｘはアミノ酸名が具体的に特定できない場合に相当する。各行と各列の交点の数値がそれらのアミノ酸の類似度であり、値が大きいほど類似度が高いことを示す。この数値はそれぞれのアミノ酸の性質等から決められている。このようなスコアテーブルは一般に数種類考案されており、検索の対象となる配列によって使い分けられている。

【００４２】ＦＡＳＴＡにおいては、検索対象のアミノ酸配列とデータベース中の配列データのうち、一方の配列中のアミノ酸をスコアテーブルの行内で探し、対応するもう一方の配列中のアミノ酸をスコアテーブルの列内で探して、それらの交点の数値をその配列データのスコアに加算する。そして、対象となる全てのアミノ酸について加算が終了した時点で、あるしきい値より大きなスコアを持つ配列データを検索対象のアミノ酸配列の類似

9

配列と考える。

【００４３】しかしながら、連続する配列のペアを順次比較するだけでは、配列データのスコアは必ずしも大きくならないので、連続する２つのアミノ酸の間にギャップを挿入して類似度を高める手法が用いられる。

【００４４】図８は、ＦＡＳＴＡにより求められた配列データの一例を示している。図８において、各アルファベットは図７のマトリクスの行に示されるようなアミノ酸の名称を表し、記号「－」はギャップを表す。これらの配列データにおいては、★印を付加した位置のアミノ酸が全て一致している。もしギャップを入れないで配列データを検索すれば、例えば下から３番目の配列Ａ３３＊

$$P=UL+V$$

（１）式においてU、Vは図６のギャップペナルティであり、この場合はU＝５、V＝２である。またLは挿入したギャップの長さ（構成単位数）である。（１）式からわかるように、ペナルティPはLの一次関数で表される。ギャップペナルティU、Vの値は、ktupとともに、ホモロジー検索のパラメータとして与えられる。これらのパラメータのとり方によってスコアが変わるため、検索される配列データも変わってくる。

【００４６】図９から図１２までは、蛋白質の一種であるCYTOCHROMEを対象としてＦＡＳＴＡにより得られたホモロジー検索の結果ファイルの一例を示している。図９にはデータベース中の検索されたファイル名が示されており、図１０には検索により類似配列として求められた配列データの個数がスコアの範囲とともに示されている。図１０において、左端の列の数値はスコアの範囲を表し、initnの列の数値はギャップを入れて計算した場合の該当するスコアを持つ配列データの個数を表し、init1 の列の数値はギャップを入れないで計算した場合の該当するスコアを持つ配列データの個数を表す。

【００４７】その右側のグラフはinitn およびinit1 の場合の配列データの個数を示している。「＝」、「－」、「＋」は、それぞれ２個分の配列データを表す。「－」はinit1 の場合の配列データを表し、「＋」はinitn の場合にギャップを入れることにより新たに増えた配列データを表す。

【００４８】図１１および図１２は、init1 のスコアが３３を越える配列データの名称とそのスコアのリストである。図１１の１行目には、４２２１５個の配列データ中の１２４１１０７６個のアミノ酸について比較が行われたことが示され、続いてinitn およびinit1 の場合のスコアの平均値が示されている。また、４行目の５８６４はinit1 の場合のスコアが３３を越える配列データの個数を表す。

【００４９】initn の列の数値はギャップを入れて計算したスコアを表し、init1 の列の数値はギャップを入れないで計算したスコアを表す。また、opt の列の数値はinit1 の場合の結果に対して公知のNeedleman-Wunsh-Se

10

＊８１３の左端には他の配列と異なるアミノ酸Ｇ、Ｓがあるため、他の配列とはかなりずれてしまい、そのスコアは小さくなる。したがって、複数の★印の位置に他の配列と一致するアミノ酸があるにも関わらず、この配列は検索結果には現れない可能性が高い。

【００４５】このようにギャップを入れることによって、見落とす可能性のあった配列データを検索することができるが、これを多用するとスコアを無制限に大きくすることができるため、類似とはいえないような配列データまで検索される危険性がある。これを防ぐために、ギャップを入れた場合には次式で算出されるペナルティＰをスコアから減じて、スコアの増大を抑えている。

（１）

llers のアライメントを行い、スコアを計算し直した値を表す。図１１および図１２においては、initn の場合のスコアの大きな順に配列名が並んでいる。検索手法や用いるパラメータが異なれば、一般に結果ファイルにおける配列名の順序が異なってくる。

【００５０】検索手法やパラメータを変えて得られる多数の結果ファイルが結果ファイル格納部１５に格納されると、ホモロジー検索結果処理プログラム１２は、得られた結果ファイル名のリストを端末１３のディスプレイ上に表示する。

【００５１】図１３は、図５に示される結果ファイルのうち、ＨＩＶに関するもののファイル名を表示した場合を示している。図１３において、ユーザは任意の複数の結果ファイル名を選択して、選択された結果ファイルの配列データに関する処理をホモロジー検索結果処理プログラム１２に行わせることができる。この選択操作は、端末１３から特定の手法名やデータベース名を入力することにより行われる。

【００５２】例えば図１３では、配列名ＨＩＶ、データベース名ＳＷ、およびパラメータ５．２．１が指定され、手法名はワイルドカード＊により無指定となっている。その結果該当するＨＩＶ_ＦＡＳＴＡ１_ＳＷ_５．２．１、ＨＩＶ ＦＡＳＴＡＮ ＳＷ ５．２．１、ＨＩＶ_ＦＡＳＴＡＯ_ＳＷ_５．２．１の３つのファイル名が網がけ表示される。ホモロジー検索結果処理プログラム１２は、これらの選択された結果ファイルを必要に応じて結果ファイル格納部１５より読み込んで、各検索結果の類似点や相違点等の特徴を抽出し、それを端末１３等に出力する。

【００５３】次に図１４から３１までを参照しながら、検索結果の特徴の表示例とその抽出方法について説明する。図１４、１７、２０、２３、２６、および２９は、図１３で選択された３つの結果ファイルについて抽出された様々な特徴の表示例を示している。これらの図において、画面上方にはTARGETとして、検索対象の配列名ＨＩＶが表示され、表示領域２１、２２、２３にはそれぞれ結果ファイルＨＩＶ ＦＡＳＴＡ１ ＳＷ５．２．

1、HIV_FASTAN_SW_5．2．1、HIV_FASTAO_SW_5．2．1に含まれる配列データのエントリー名がスコアの高い順に表示されている。この場合はLOCUS 1 等が1つの配列データに相当するエントリー名である。また、表示領域21、22、23内の上部にはそれぞれの結果ファイルの手法名、パラメータ、データベース名が表示されている。

【0054】また、図15、16、18、19、21、22、24、25、27、28、30、31は、ホモロジー検索結果処理プログラム12による特徴抽出処理のフローチャートである。

【0055】図14は、選択された結果ファイルの全てに含まれている配列の表示例を示している。図14において、エントリー名LOCUS 1 、LOCUS 2 、LOCUS 3 、LOCUS5 は3つの結果ファイルの全てに含まれているので、結果ファイル間の類似点として網がけ表示される。

【0056】図15は、結果ファイルの選択処理および図14の類似点を抽出する処理のフローチャートである。図15において処理が開始されると、例えば図13のような結果ファイル名のリストが表示される（ステップS1）。次にユーザにより端末13から結果ファイル名の指定情報が入力されると（ステップS2）、指定情報に該当する結果ファイル名が選択され、網がけ表示される（ステップS3）。続いて、選択された各結果ファイルの内容が例えば図14のように表示される。このとき選択された結果ファイルの数はjmaxとして不図示のメモリに記憶される。次にユーザが、選択された結果ファイルに共通して含まれるエントリー名の検索を指示すると（ステップS4）、指示された検索が行われ、その結果が網がけ表示される（ステップS5）。

【0057】図16は、ホモロジー検索結果処理プログラム12による図15のステップS5の処理のフローチャートである。図16の処理は、図15の処理から呼び出されるサブルーチン、あるいは図15の処理とは別のプロセスにより実行される。図16において、まず各結果ファイル内の配列のエントリー名をサブルーチンあるいは別プロセスに入力し（ステップS11）、i＝1、1番目の結果ファイルのエントリー数をimaxとおく（ステップS12）。例えば図14の場合は、1番目の結果ファイルの内容は表示領域21に表示されており、imax＝6となる。また2番目、3番目の結果ファイルの内容は、それぞれ表示領域22、23に表示されている。

【0058】次にj＝2とおき（ステップS13）、1番目の結果ファイルのi番目のエントリー名がj番目の結果ファイルに含まれるかどうかを判定する（ステップS14）。判定結果がYESの場合は、jに1を加算し（ステップS15）、続いてjとjmaxの値を比較する（ステップS16）。ステップS16でjがjmaxを越えていなければ（ステップS16、YES）、ステップS14以降の処理を繰り返す。

【0059】ステップS16でjがjmaxを越えると（ステップS16、NO）、そのエントリー名が全ての結果ファイルに含まれることが分かるので、全ての結果ファイルの該当するエントリー名にフラグ（FLAG）を立てる（ステップS17）。そしてiに1を加算し（ステップS18）、続いてiとimaxの値を比較する（ステップS19）。

【0060】ステップS14で判定結果がNOの場合は、ステップS15からS17の処理を行わずに、ステップS18の処理に進む。ステップS19でiがimaxを越えていなければ（ステップS19、YES）、ステップS13以降の処理を繰り返し、iがimaxを越えると処理を終了する（ステップS19、NO）。その後、図15のステップS5では、フラグの立っている各結果ファイルのエントリー名が網がけ表示される。

【0061】例えば図14の場合は、i＝1、2、3、5にそれぞれ対応するLOCUS 1 、LOCUS 2 、LOCUS 3 、LOCUS 5 の各エントリー名についてステップS17の処理が行われ、表示領域21、22、23内のこれらのエントリー名が網がけ表示される。i＝4、6に対応するエントリー名LOCUS 4 、LOCUS 6 については、それぞれj＝3、2のときにステップS14の判定結果がNOとなり、ステップS17の処理は行われない。したがって、これらのエントリー名は網がけ表示されない。

【0062】図17は、指定した結果ファイルの中の配列とは異なる配列の表示例を示している。図17においては、ユーザにより表示領域22の結果ファイルが指定され、他の結果ファイル内のエントリー名のうちLOCUS 6 とLOCUS 8 が指定された結果ファイルに含まれないので、これらのエントリー名が指定された結果ファイルとの相違点として表示領域21、22内で網がけ表示される。

【0063】図18は、結果ファイルの選択処理および図17の相違点を抽出する処理のフローチャートである。図18のステップS21、S22、S23の処理は、図15のステップS1、S2、S3の処理と同様である。ステップS23の処理に続いて、ユーザが結果ファイルの指定を行い、その結果ファイルに含まれないエントリー名の検索を指示すると（ステップS24）、指示された検索が行われ、その結果が網がけ表示される（ステップS25）。

【0064】図19は、ホモロジー検索結果処理プログラム12による図17のステップS25の処理のフローチャートである。図19において、まず指定された結果ファイルを1番目として、各結果ファイル内の配列のエントリー名をサブルーチン等に入力し（ステップS31）、i＝1、1番目の結果ファイルのエントリー数をimaxとおく（ステップS32）。例えば図17の場合は、1番目の結果ファイルの内容は表示領域22に表

示されており、ｉｍａｘ＝６である。また２番目、３番目の結果ファイルの内容は、それぞれ表示領域２１、２３に表示されている。

【００６５】次にｊ＝２とおき（ステップＳ３３）、指定された１番目の結果ファイルのｉ番目のエントリー名がｊ番目の結果ファイルに含まれるかどうかを判定する（ステップＳ３４）。判定結果がＹＥＳの場合は、ｊ番目の結果ファイルのそのエントリー名に負のフラグを立て（ステップＳ３５）、ｊに１を加算し（ステップＳ３６）、続いてｊとｊｍａｘの値を比較する（ステップＳ３７）。ステップＳ３４で判定結果がＮＯの場合は、ステップＳ３５の処理を行わずにステップＳ３６の処理に進む。

【００６６】ステップＳ３７でｊがｊｍａｘを越えていなければ（ステップＳ３７、ＹＥＳ）、ステップＳ３４以降の処理を繰り返す。ステップＳ３７でｊがｊｍａｘを越えると（ステップＳ３７、ＮＯ）、ｉに１を加算し（ステップＳ３８）、続いてｉとｉｍａｘの値を比較する（ステップＳ３９）。

【００６７】ステップＳ３９でｉがｉｍａｘを越えていなければ（ステップＳ３９、ＹＥＳ）、ステップＳ３３以降の処理を繰り返し、ｉがｉｍａｘを越えると処理を終了する（ステップＳ３９、ＮＯ）。その後、図１８のステップＳ２５では、負のフラグが立っていない、指定されなかった結果ファイルのエントリー名が網がけ表示される。

【００６８】例えば図１７の場合は、指定された表示領域２２の結果ファイルに含まれるLOCUS １、LOCUS ２、LOCUS ４、LOCUS ７、LOCUS ３、LOCUS ５の各エントリー名についてステップＳ３５の処理が行われ、表示領域２１、２３内のこれらのエントリー名は網がけ表示されない。一方、表示領域２１内のLOCUS ６と表示領域２３内のLOCUS ８についてはステップＳ３５の処理は行われないので、これらのエントリー名が網がけ表示される。

【００６９】図２０は、選択された結果ファイルの全てに含まれている配列とは異なる配列の表示例を示している。選択された結果ファイルの全てに含まれている配列のエントリー名は、図１４に示したとおり、LOCUS １、LOCUS ２、LOCUS ３、LOCUS5 の４つである。図２０においては、それら以外のエントリー名であるLOCUS ４、LOCUS ６、LOCUS ７、LOCUS ８が、結果ファイル間の相違点として各表示領域内で網がけ表示される。

【００７０】図２１は、結果ファイルの選択処理および図２０の相違点を抽出する処理のフローチャートである。図２１のステップＳ４１、Ｓ４２、Ｓ４３の処理は、図１５のステップＳ１、Ｓ２、Ｓ３の処理と同様である。ステップＳ４３の処理に続いて、ユーザが選択された結果ファイルの全てに含まれている配列とは異なる配列の検索を指示すると（ステップＳ４４）、指示された検索が行われ、その結果が網がけ表示される（ステップＳ４５）。

【００７１】図２２は、ホモロジー検索結果処理プログラム１２による図２１のステップＳ４５の処理のフローチャートである。図２２において、まず各結果ファイル内の配列のエントリー名をサブルーチン等に入力し（ステップＳ５１）、ｉ＝１、１番目の結果ファイルのエントリー数をｉｍａｘとおく（ステップＳ５２）。例えば図２０の場合は、１番目の結果ファイルの内容は表示領域２１に表示されており、ｉｍａｘ＝６である。また２番目、３番目の結果ファイルの内容は、それぞれ表示領域２２、２３に表示されている。

【００７２】次にｊ＝２とおき（ステップＳ５３）、１番目の結果ファイルのｉ番目のエントリー名がｊ番目の結果ファイルに含まれるかどうかを判定する（ステップＳ５４）。判定結果がＹＥＳの場合は、ｊに１を加算し（ステップＳ５５）、続いてｊとｊｍａｘの値を比較する（ステップＳ５６）。ステップＳ５６でｊがｊｍａｘを越えていなければ（ステップＳ５６、ＹＥＳ）、ステップＳ５４以降の処理を繰り返す。

【００７３】ステップＳ５６でｊがｊｍａｘを越えると（ステップＳ５６、ＮＯ）、そのエントリー名が全ての結果ファイルに含まれることが分かるので、これを網がけ表示させないために全ての結果ファイルの該当するエントリー名に負のフラグを立てる（ステップＳ５７）。そしてｉに１を加算し（ステップＳ５８）、続いてｉとｉｍａｘの値を比較する（ステップＳ５９）。

【００７４】ステップＳ５４で判定結果がＮＯの場合は、ステップＳ５５からＳ５７の処理を行わずに、ステップＳ５８の処理に進む。ステップＳ５９でｉがｉｍａｘを越えていなければ（ステップＳ５９、ＹＥＳ）、ステップＳ５３以降の処理を繰り返し、ｉがｉｍａｘを越えると処理を終了する（ステップＳ５９、ＮＯ）。その後、図２１のステップＳ４５では、負のフラグが立っていない、全ての結果ファイルのエントリー名が網がけ表示される。

【００７５】例えば図２０の場合は、ｉ＝１、２、３、５にそれぞれ対応するLOCUS １、LOCUS ２、LOCUS ３、LOCUS ５の各エントリー名についてステップＳ５７の処理が行われ、表示領域２１、２２、２３内のこれらのエントリー名は網がけ表示されない。ｉ＝４、６に対応するエントリー名LOCUS ４、LOCUS ６については、それぞれｊ＝３、２のときにステップＳ５４の判定結果がＮＯとなり、ステップＳ５７の処理は行われない。また、１番目の結果ファイルに含まれないエントリー名LOCUS ７、LOCUS ８についてもステップＳ５７の処理は行われない。したがって、LOCUS ４、LOCUS ６、LOCUS ７、LOCUS ８のみが網がけ表示される。

【００７６】図２３は、指定した配列が結果ファイルの中に含まれているかどうかを表示した例を示している。図２３においては、ユーザによりエントリー名LOCUS ４

15

とLOCUS 5 が指定され、それぞれ別の網目を用いて網がけ表示される。このように、ユーザは任意の配列を指定して、それが選択された結果ファイル内に含まれているか否か、また、どの結果ファイル内に含まれているかを認識することができる。

【0077】図24は、結果ファイルの選択処理および図23に示された指定エントリー名を抽出する処理のフローチャートである。図24のステップS61、S62、S63の処理は、図15のステップS1、S2、S3の処理と同様である。ステップS63の処理に続いて、ユーザが特定の配列のエントリー名を指定して、選択された結果ファイル内でそのエントリー名の検索を指示すると（ステップS64）、指示された検索が行われ、その結果が網がけ表示される（ステップS65）。

【0078】図25は、ホモロジー検索結果処理プログラム12による図24のステップS65の処理のフローチャートである。図25において、まず各結果ファイル内の配列のエントリー名をサブルーチン等に入力し（ステップS71）、j＝1とおく（ステップS72）。

【0079】次にi＝1、j番目の結果ファイルのエントリー数をimaxとおく（ステップS73）。例えば図23の場合は、表示領域21、22、23の結果ファイルをそれぞれ1、2、3番目の結果ファイルとする。これらの結果ファイルのエントリー数はいずれも6なので、imaxは6となる。

【0080】次にj番目の結果ファイルのi番目のエントリー名が指定されたエントリー名かどうかを判定する（ステップS74）。判定結果がYESの場合は、j番目の結果ファイルのi番目のエントリー名にフラグを立てる（ステップS77）。そしてjに1を加算し（ステップS78）、続いてjとjmaxの値を比較する（ステップS59）。

【0081】ステップS74で判定結果がNOの場合は、iに1を加算し（ステップS75）、続いてiとimaxの値を比較する（ステップS76）。ステップS76でiがimaxを越えていなければ（ステップS76、YES）、ステップS74以降の処理を繰り返す。ステップS76でiがimaxを越えると（ステップS76、NO）、ステップS78の処理に進む。

【0082】ステップS79でjがjmaxを越えていなければ（ステップS79、YES）、ステップS73以降の処理を繰り返し、jがjmaxを越えると処理を終了する（ステップS79、NO）。その後、図24のステップS65では、フラグの立っている各結果ファイルのエントリー名が網がけ表示される。

【0083】例えば図23の場合は、ユーザにより最初にエントリー名LOCUS 4 が指定され、j＝1、i＝4のときとj＝2、i＝3のときにステップS77の処理が行われ、表示領域21、22内のエントリー名LOCUS 4 が網がけ表示される。次にエントリー名LOCUS 5 が指定

16

され、j＝1、i＝5のときとj＝2、i＝6のとき、およびj＝3、i＝4のときにステップS77の処理が行われ、表示領域21、22、23内のエントリー名LOCUS 5 が、別の網目により網がけ表示される。他のエントリー名については、ステップS74の判定結果がNOとなり、ステップS77の処理は行われないので、網がけ表示されない。

【0084】図26は、選択された結果ファイルのデータを重ねて、一致している部分を類似点として強調表示した例を示している。図26（a）においては、エントリー名LOCUS 1 とLOCUS 2 が全ての表示領域の1番目に表示されており、LOCUS 3 は表示領域21と23の2番目に表示されている。したがって、ユーザが重ね合わせ表示を指示すると、図26（b）に示すようにLOCUS 1 とLOCUS 2 が同じ網目を用いて網がけ表示され、LOCUS 3 は別の網目で表示される。他のエントリー名については各結果ファイルの間で重なりが生じていないので、網がけ表示されずに個別に表示される。重なりが多いエントリー名ほどその順位の信頼性が高いので、これにより得られた検索結果の相同性の信頼度が示される。

【0085】図27は、結果ファイルの選択処理および図26（b）の類似点を抽出する処理のフローチャートである。図27のステップS81、S82、S83の処理は、図15のステップS1、S2、S3の処理と同様である。ステップS83の処理に続いて、ユーザが重ね合わせ表示を指示すると（ステップS84）、結果ファイルのリスト内において同順位の同じエントリー名の数が求められ（ステップS85）、その数に応じてエントリー名が網がけ表示される（ステップS86）。

【0086】図28は、ホモロジー検索結果処理プログラム12による図27のステップS85の処理のフローチャートである。図28において、まず各結果ファイル内の配列のエントリー名をサブルーチン等に入力し（ステップS91）、i＝1、結果ファイルのエントリー数の最大値をimaxとおく（ステップS92）。例えば図26（a）の場合は、結果ファイルのエントリー数はいずれも6なので、imax＝6である。また、表示領域21、22、23に表示された結果ファイルをそれぞれ順に1、2、3番目の結果ファイルとする。

【0087】次にj＝1、n＝1、nmax＝1、k（1）＝1、k（n）＝0（n＝2，・・・，jmax）とおき（ステップS93）、j番目の結果ファイルのi番目のエントリー名をn番目の検索対象にする（ステップS94）。次にj＝j＋1とおき（ステップS95）、n番目の検索対象のエントリー名がj番目の結果ファイルのi番目のエントリー名と一致するか否かを判定する（ステップS96）。判定結果がNOの場合は、n＝n＋1とおき（ステップS97）、nとnmaxの値を比較する（ステップS98）。nがnmaxを越えていなければ（ステップS98、NO）、ステップS9

６以降の処理を繰り返す。

【００８８】ステップＳ９８でｎがｎｍａｘを越えれば（ステップＳ９８、ＹＥＳ）、ｊ番目の結果ファイルのｉ番目のエントリー名をｎ番目の検索対象にする（ステップＳ９９）。続いてｎｍａｘに１を加算し（ステップＳ１００）、ｋ（ｎ）に１を加算し（ステップＳ１０１）、ｊに１を加算してｎ＝１とおいて（ステップＳ１０２）、ｊとｊｍａｘの値を比較する（ステップＳ１０３）。

【００８９】ステップＳ９６で判定結果がＹＥＳの場合は、ステップＳ９７～Ｓ１００の処理を行わずに、ステップＳ１０１以降の処理を行う。ステップＳ１０３でｊがｊｍａｘを越えていなければ（ステップＳ１０３、ＹＥＳ）、ステップＳ９６以降の処理を繰り返し、ｊがｊｍａｘを越えると（ステップＳ１０３、ＮＯ）、ｉに１を加算して（ステップＳ１０４）、続いてｉとｉｍａｘの値を比較する（ステップＳ１０５）。

【００９０】ステップＳ１０５でｉがｉｍａｘを越えていなければ（ステップＳ１０５、ＹＥＳ）、ステップＳ９３以降の処理を繰り返し、ｉがｉｍａｘを越えると処理を終了する（ステップＳ１０５、ＮＯ）。その後、図２７のステップＳ８６では、スコアの順位毎にｎ番目の検索対象に指定されたエントリー名が左から順に表示され、ｋ（ｎ）の値に応じて異なる網目を用いて網がけ表示される。どの網目を用いるかは環境変数により決められている。

【００９１】例えば図２６（ａ）の場合は、ｉ＝１のとき３つの結果ファイルには同じエントリー名LOCUS １が含まれるのでｎ＝１についてのみ検索が行われ、ｊ＝２、３についてステップＳ９６の判定結果がＹＥＳとなる。したがって、LOCUS １に対応するｋ（１）は２回インクリメントされて（ステップＳ１０１）、３になる。ｉ＝２のときも同様である。

【００９２】ｉ＝３のとき、まず１番目の結果ファイルの３番目のエントリー名LOCUS 3 が１番目の検索対象となり（ステップＳ９４）、LOCUS 3 が２番目の結果ファイルの３番目にないので（ステップＳ９６、ＮＯ）、次に２番目の結果ファイルの３番目のエントリー名LOCUS 4 が２番目の検索対象となる（ステップＳ９９）。そしてLOCUS 4 に対応するｋ（２）がインクリメントされて１になる（ステップＳ１０１）。次にLOCUS 3 が３番目の結果ファイルの３番目にあるので（ステップＳ９６、ＹＥＳ）、LOCUS 3 に対応するｋ（１）がインクリメントされて２になる（ステップＳ１０１）。ここで、ｊ＝４＞３＝ｊｍａｘとなるため（ステップＳ１０３、ＮＯ）、ｉがインクリメントされる。

【００９３】ｉ＝４のとき、まず１番目の結果ファイルの４番目のエントリー名LOCUS 4 が１番目の検索対象となり（ステップＳ９４）、LOCUS 4 が２番目の結果ファイルの４番目にないので（ステップＳ９６、ＮＯ）、次に２番目の結果ファイルの４番目のエントリー名LOCUS 7 が２番目の検索対象となる（ステップＳ９９）。そしてLOCUS 7 に対応するｋ（２）がインクリメントされて１になる（ステップＳ１０１）。

【００９４】ところが、LOCUS 4 とLOCUS 7 のいずれも３番目の結果ファイルの４番目にないので（ステップＳ９６、ＮＯ）、３番目の結果ファイルの４番目のエントリー名LOCUS 5 が３番目の検索対象となり（ステップＳ９９）、LOCUS 5 に対応するｋ（３）がインクリメントされて１になる（ステップＳ１０１）。ここで、ｊ＝４になるため（ステップＳ１０２）、ｉがインクリメントされる。LOCUS 5 に対応するｋ（１）はインクリメントされず、１のままである（ステップＳ９３）。ｉ＝５、６のときも同様である。

【００９５】図２７のステップＳ８６では、こうして得られた各ｋ（ｎ）の値毎に異なる網目を用いて、図２６（ｂ）に示すように対応するエントリー名が網がけ表示される。ただし、ｋ（ｎ）＝１のエントリー名については強調する必要がないので、網がけ表示されない。

【００９６】図２９は、選択された結果ファイル内の配列の数をグラフ表示した例を示している。図２９（ａ）においてユーザがグラフ表示を指示すると、図２９（ｂ）に示すように、全ての結果ファイルに含まれる各エントリー名の個数Ｎがグラフ表示される。結果ファイルに多く含まれるエントリー名の配列は検索対象の配列との類似度が高いと考えられ、グラフ化することにより検索結果の信頼性が示される。例えばLOCUS ４、LOCUS ５、LOCUS 7 は、図２６（ｂ）の重ね合わせ表示においては強調されていないが、表示領域２１、２２、２３に複数現れていることが明らかになる。

【００９７】図３０は、結果ファイルの選択処理および図２９（ｂ）のグラフ化処理のフローチャートである。図３０のステップＳ１１１、Ｓ１１２、Ｓ１１３の処理は、図１５のステップＳ１、Ｓ２、Ｓ３の処理と同様である。ステップＳ１１３の処理に続いて、ユーザがグラフ表示を指示すると（ステップＳ１１４）、結果ファイル内の同じエントリー名の総数が求められ（ステップＳ１１５）、その数がエントリー名毎にグラフ表示される（ステップＳ１１６）。

【００９８】図３１は、ホモロジー検索結果処理プログラム１２による図３０のステップＳ１１５の処理のフローチャートである。図３１において、まず各結果ファイル内の配列のエントリー名をサブルーチン等に入力し（ステップＳ１２１）、結果ファイルのエントリー数の最大値をｉｍａｘ、ｎ＝１とおく（ステップＳ１２２）。例えば図２９（ａ）の場合はｉｍａｘ＝６である。また、表示領域２１、２２、２３に表示された結果ファイルをそれぞれ順に１、２、３番目の結果ファイルとする。

【００９９】次にｉ＝１とおき（ステップＳ１２３）、

19

ｎ番目の結果ファイルのｉ番目のエントリー名にフラグが立っているか否かを判定する。判定結果がＮＯであればｊ＝ｎ＋１、ｋ＝１とおき（ステップＳ１２５）、続いてｎ番目の結果ファイルのｉ番目のエントリー名がｊ番目の結果ファイルに含まれるか否かを判定する（ステップＳ１２６）。この判定結果がＹＥＳの場合は、ｋ＝ｋ＋１とおき（ステップＳ１２７）、ｊ番目の結果ファイルのｉ番目のエントリー名にフラグを立てる（ステップＳ１２８）。次にｊに１を加算し（ステップＳ１２９）、ｊとｊｍａｘの値を比較する（ステップＳ１３０）。

【０１００】ステップＳ１２６で判定結果がＮＯの場合は、ステップＳ１２９以降の処理を行う。ステップＳ１３０でｊがｊｍａｘを越えていなければ（ステップＳ１３０、ＹＥＳ）、ステップＳ１２６以降の処理を繰り返し、ｊがｊｍａｘを越えると（ステップＳ１３０、ＮＯ）、ｎ番目の結果ファイルのｉ番目のエントリー名をｋの値とともに不図示のメモリに格納する（ステップＳ１３１）。続いてｉに１を加算し（ステップＳ１３２）、ｉとｉｍａｘの値を比較する（ステップＳ１３３）。ステップＳ１２４で判定結果がＹＥＳの場合は、ステップＳ１３２以降の処理を行う。

【０１０１】ステップＳ１３３でｉがｉｍａｘを越えていなければ（ステップＳ１３３、ＹＥＳ）、ステップＳ１２４以降の処理を繰り返す。ｉがｉｍａｘを越えるとｎに１を加算し（ステップＳ１３４）、ｎとｊｍａｘの値を比較する（ステップＳ１３５）。

【０１０２】ステップＳ１３５でｎがｊｍａｘを越えていなければ（ステップＳ１３５、ＹＥＳ）、ステップＳ１２３以降の処理を繰り返し、ｎがｊｍａｘを越えると処理を終了する（ステップＳ１３５、ＮＯ）。その後、図３０のステップＳ１１６では、上記メモリに格納された各エントリー名とそのｋの値が順に取り出され、このｋの値がそのエントリー名の個数としてグラフ表示される。

【０１０３】例えば図２９（ａ）の場合は、ｎ＝１、ｉ＝１のとき、１番目の結果ファイルの１番目のエントリー名LOCUS １が、２番目、３番目の結果ファイルにも含まれているので（ステップＳ１２６、ＹＥＳ）、ｊ＝２、３についてｋがインクリメントされ（ステップＳ１２７）、２番目、３番目の結果ファイルのLOCUS １にフラグが立てられる（ステップＳ１２８）。このときｋ＝３がエントリー名LOCUS1 の個数として記憶される（ステップＳ１３１）。ｉ＝２、３、５にそれぞれ相当するLOCUS ２、LOCUS ３、LOCUS ５についても同様である。

【０１０４】次にｉ＝４のとき、１番目の結果ファイルの４番目のエントリー名LOCUS ４は２番目の結果ファイルには含まれているが（ステップＳ１２６、ＹＥＳ）、３番目の結果ファイルには含まれていないので（ステップＳ１２６、ＮＯ）、ｊ＝２についてのみｋがインクリメントされ（ステップＳ１２７）、２番目の結果ファイルのLOCUS ４にフラグが立てられる（ステップＳ１２８）。このときｋ＝２がエントリー名LOCUS ４の個数として記憶される（ステップＳ１３１）。

20

【０１０５】次にｉ＝６のとき、１番目の結果ファイルの６番目のエントリー名LOCUS ６は他の結果ファイルに含まれていないので（ステップＳ１２６、ＮＯ）、ｋはインクリメントされず、ｋ＝１がエントリー名LOCUS ６の個数として記憶される（ステップＳ１３１）。

【０１０６】次にｉをインクリメントすると（ステップＳ１３２）、ｉがｉｍａｘを越えるので（ステップＳ１３３、ＮＯ）、ｎ＝２として（ステップＳ１３４）ｉを初期化し（ステップＳ１２３）、同様の処理を行う。ここでｉ＝１、２、３、５、６のときは、そのエントリー名にフラグが立っているため（ステップＳ１２４、ＹＥＳ）、ｋはインクリメントされずステップＳ１３１の処理も行われない。

【０１０７】ｉ＝４のとき、２番目の結果ファイルの４番目のエントリー名LOCUS ７にはフラグが立っていなくて（ステップＳ１２４、ＮＯ）、しかもLOCUS ７は３番目の結果ファイルに含まれるので（ステップＳ１２６、ＹＥＳ）、ｊ＝３についてｋがインクリメントされ（ステップＳ１２７）、３番目の結果ファイルのLOCUS ７にフラグが立てられる（ステップＳ１２８）。このときｋ＝２がエントリー名LOCUS ７の個数として記憶される（ステップＳ１３１）。

【０１０８】次にｎ＝３のときは、ｉ＝６のエントリー名LOCUS ８についてのみステップＳ１２６の判定が行われるが、４番目の結果ファイルは存在しないので（ステップＳ１２６、ＮＯ）、ｋはインクリメントされず、ｋ＝１がエントリー名LOCUS ８の個数として記憶される（ステップＳ１３１）。

【０１０９】こうして記憶された各エントリー名の個数が、図２９（ｂ）のようなグラフとして表示される（ステップＳ１１６）。以上の実施例では、選択された複数の結果ファイルについてのいくつかの特徴抽出の例を示したが、本発明はこれらに限定されることはなく、ホモロジー検索結果処理プログラム１２として他のアプリケーションを用意することにより、任意の他の特徴を抽出する構成とすることもできる。

【０１１０】また、図１３、１４等において、ファイル名やエントリー名を網がけ表示しているが、多色あるいは多階調のマーカーを用いてエントリー名をマークして表示する構成にしてもよい。本発明においては、結果ファイル内の検索結果をその後の結果の比較やアライメント等の解析のデータとして利用するために、選択により結果ファイル全体を保存したり、マークされた配列の情報のみを保存したり、あるいはマークされなかった配列の情報のみを保存したりすることができる。

【０１１１】さらに、本発明は、バイオテクノロジー分

野におけるホモロジー検索結果の処理に限らず、順位付けられたデータ項目を有する複数のリストの間の任意の特徴を抽出する処理に適用することができる。

【０１１２】

【発明の効果】本発明によれば、ホモロジー検索結果に見られるような、互いに類似点と相違点を有する複数のリストを効率よく比較することができる。ホモロジー検索結果の場合、膨大な数の配列名を含む多数のリストの比較を迅速に行うことができる。

【０１１３】複数のリストを比較してそれらの類似点を抽出することにより、信頼性の高いデータが得られる。ホモロジー検索結果においては、多数のリストに共通して含まれる配列名等の良質なデータを効率よく抽出することができる。

【０１１４】また、抽出された各種の特徴が分かりやすく画面表示されるので、それらの特徴の把握が容易になり、それらを選択して保存することにより、他のシステムや装置へのデータの受け渡しも容易になる。

【０１１５】さらに、ホモロジー検索結果を格納するファイルを、使用した検索手法等の項目名を用いて管理しているので、多数のファイルの中から特定の項目名やパラメータを持つファイルを選択する操作が容易になる。

【図面の簡単な説明】

【図１】本発明の原理図である。

【図２】ホモロジー検索結果に影響を与える項目を示す図である。

【図３】本発明の実施例の構成図である。

【図４】本発明の実施例における結果ファイルの命名規約を示す図である。

【図５】本発明の実施例における結果ファイル名を示す図である。

【図６】本発明の実施例における結果ファイルのデータ構造を示す図である。

【図７】ホモロジー検索において用いられるスコアテーブルの一例を示す図である。

【図８】検索された配列データを示す図である。

【図９】ＦＡＳＴＡによる結果ファイルの一例を示す図（その１）である。

【図１０】ＦＡＳＴＡによる結果ファイルの一例を示す図（その２）である。

【図１１】ＦＡＳＴＡによる結果ファイルの一例を示す図（その３）である。

【図１２】ＦＡＳＴＡによる結果ファイルの一例を示す図（その４）である。

【図１３】本発明の実施例における結果ファイル名の画面表示を示す図である。

【図１４】本発明の実施例における共通する配列名の画面表示を示す図である。

【図１５】本発明の実施例における共通する配列名の表示処理を示すフローチャートである。

【図１６】本発明の実施例における共通する配列名の抽出処理を示すフローチャートである。

【図１７】本発明の実施例における指定された結果ファイルに含まれない配列名の画面表示を示す図である。

【図１８】本発明の実施例における指定された結果ファイルに含まれない配列名の表示処理を示すフローチャートである。

【図１９】本発明の実施例における指定された結果ファイルに含まれない配列名の抽出処理を示すフローチャートである。

【図２０】本発明の実施例における共通して含まれない配列名の画面表示を示す図である。

【図２１】本発明の実施例における共通して含まれない配列名の表示処理を示すフローチャートである。

【図２２】本発明の実施例における共通して含まれない配列名の抽出処理を示すフローチャートである。

【図２３】本発明の実施例における指定された配列名の画面表示を示す図である。

【図２４】本発明の実施例における指定された配列名の表示処理を示すフローチャートである。

【図２５】本発明の実施例における指定された配列名の抽出処理を示すフローチャートである。

【図２６】本発明の実施例における共通する同順位の配列名の画面表示を示す図である。

【図２７】本発明の実施例における共通する同順位の配列名の表示処理を示すフローチャートである。

【図２８】本発明の実施例における共通する同順位の配列名の抽出処理を示すフローチャートである。

【図２９】本発明の実施例における配列名の数のグラフ表示を示す図である。

【図３０】本発明の実施例における配列名の数のグラフ表示処理を示すフローチャートである。

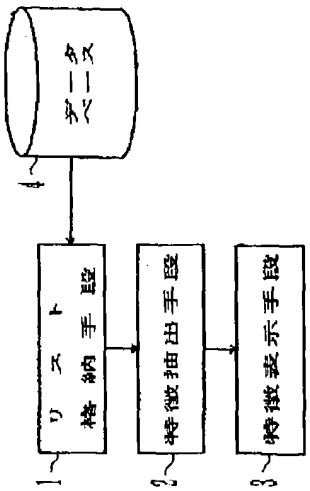
【図３１】本発明の実施例における配列名の数の計算処理を示すフローチャートである。

【符号の説明】

１　リスト格納手段
２　特徴抽出手段
３　特徴表示手段
４　データベース

【図１】

本発明の原理図



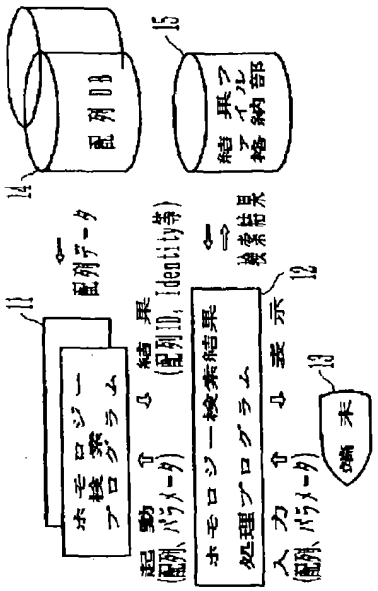
【図２】

ホモロジー検索における分類項目を示す図

分類項目	内容
配列	検索の対象となる遺伝子またはアミノ酸配列
ホモロジー検索手法	FASTA, BLAST, Smith-Watermanの検索手法
データベース	GenBank, EMBL, DDBJ, NRRF, SWISS-PROT等の配列データベース
パラメータ	ホモロジー検索手法のパラメータで結果に影響を与える物

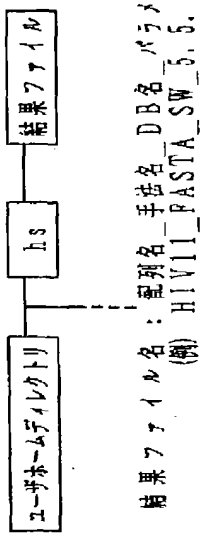
【図３】

実施例の構成図



【図４】

結果ファイルの命名規約を示す図



【図９】

FASTAによる結果ファイルを示す図（その１）

```
>2CPP  CYTOCHROME P450CAM (CAMPHOR MONOOXYGENASE) : 405 aa
vs pir library
 search file :
    /rie/homel/pir33/pir1.dat
    /rie/homel/pir33/pir2.dat
    /rie/homel/pir33/pir3.dat
using protein matrix
```

【図１１】

FASTAによる結果ファイルを示す図（その３）

```
12411016 residues in 42215 sequences
   mean initn score:  26.0 (7.35)
   mean init1 score:  25.5 (6.51)
   5864 scores better than 33 saved, ktup: 2, fact: 8  scan time:  0:02:25
The best scores are:                                      initn init1   opt
>A25660  Cytochrome P450 101 - Pseudomonas putida  2050  2050  2050
>O46SCP  Cytochrome P450 101 - Pseudomonas putida  1999  1801  2003
>O46526  Cytochrome P450 126 - Bacillus megaterium  257    99   364
>AM0401  Cytochrome P450 55 precursor - Imperfect   216   131   282
>S18324  *cytochrome P-450gcy - Streptomyces grise  172   111   332
>A35401  Cytochrome P450 105A1 - Streptomyces griс  160   121   393
>S35401  Cytochrome P450 105B1 - Streptomyces griс  155    93   367
>S15809  *cytochrome P450-like protein - Streptomy  143    90   325
>A27555  Steroid 21-monooxygenase cytochrome P450   111    76    78
>A24101  Steroid 21-monooxygenase cytochrome P450   111    74    76
```

【図5】

結果ファイル名を示す図

```
HIV__FASTA1__SW__5.2.1      LYSO __FASTA1__SW__5.2.1      UBIQ __FASTA1__SW__5.2.1
HIV__FASTA1__SW__5.5.1      LYSO __FASTAN__SW__5.2.1      UBIQ __FASTAN__SW__5.2.1
HIV__FASTAN__SW__5.2.1      LYSO __FASTA1__SW__5.5.1      UBIQ __FASTA1__SW__5.5.1
HIV__FASTAN__SW__5.5.1      LYSO __FASTAN__SW__5.5.1      UBIQ __FASTAN__SW__5.5.1
HIV__FASTA0__SW__5.2.1                 .                             .
HIV__FASTA0__SW__5.5.1                 .                             .
HIV__SW-WT __SW__5.2.2                 .                             .
HIV__SW-WT __SW__5.5.1      LECTIN __FASTA1__SW__5.2.1    TRYPSIN__FASTA1__SW__5.2.1
       .                    LECTIN __FASTAN__SW__5.2.1    TRYPSIN__FASTAN__SW__5.2.1
       .                    LECTIN __FASTA1__SW__5.5.1    TRYPSIN__FASTA1__SW__5.5.1
       .                    LECTIN __FASTAN__SW__5.5.1    TRYPSIN__FASTAN__SW__5.5.1
```

【図6】

結果ファイルのデータ構造を示す図

```
Param：ホモロジー検索の手法で用いられるパラメータの詳細情報
      （例） Gap Penalty U= 5, Gap Penalty V= 2, Ktup=1
Sequence：検索対象(TARGET)配列の名前
      （例） HIV-1 PROTEASE
DB：検索を行うデータベースの名前
      （例） SWISS-PROT
LIST：検索された配列の名前（各データベースごと付けられている配列のエントリー名）
    及びスコア（配列の類似度を示す）
    （例） HIV __MANMA     1133
           HIV __YEAST     1127
           HIV __THEPS     1018
           HIV __DESVH      999
           HIV __ECOLI      986
           HIV __SCOPD      503
           HIV __ARATH      198
           HIV __PEAAU      174
              .              .
              .              .
```

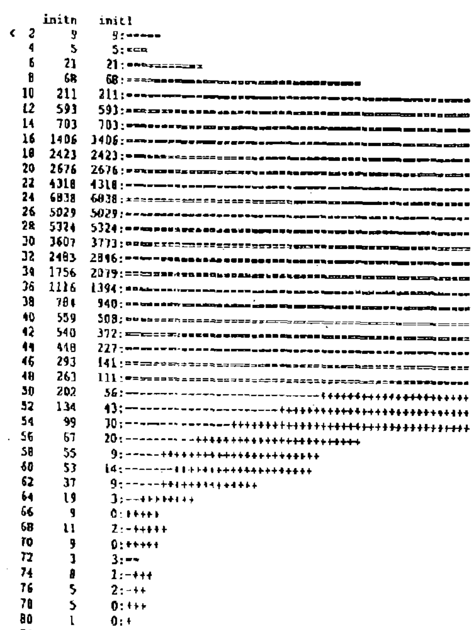
【図7】

スコアテーブルの一例を示す図

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	Asx	Glx	???
A Ala	2																						
R Arg	-2	6																					
N Asn	0	0	2																				
D Asp	0	-1	2	4																			
C Cys	-2	-4	-4	-5	12																		
Q Gln	0	1	1	2	-5	4																	
E Glu	0	-1	1	3	-5	2	4																
G Gly	1	-3	0	1	-3	-1	0	5															
H His	-1	2	2	1	-3	3	1	-2	6														
I Ile	-1	-2	-2	-2	-2	-2	-2	-3	-2	5													
L Leu	-2	-3	-3	-4	-6	-2	-3	-4	-2	2	6												
K Lys	-1	3	1	0	-5	1	0	-2	0	-2	-3	5											
M Met	-1	0	-2	-3	-5	-1	-2	-3	-2	2	4	0	6										
F Phe	-4	-4	-4	-6	-4	-5	-5	-5	-2	1	2	-5	0	9									
P Pro	1	0	-1	-1	-3	0	-1	-1	0	-2	-3	-1	-2	-5	6								
S Ser	1	0	1	0	0	-1	0	1	-1	-1	-3	0	-2	-3	1	2							
T Thr	1	-1	0	0	-2	-1	0	0	-1	0	-2	0	-1	-3	0	1	3						
W Trp	-6	2	-4	-7	-8	-5	-7	-7	-3	-5	-2	-3	-4	0	-6	-2	-5	17					
Y Tyr	-3	-4	-2	-4	0	-4	-4	-5	0	-1	-1	-4	-2	7	-5	-3	-3	0	10				
V Val	0	-2	-2	-2	-2	-2	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	4			
B Asx	0	-1	2	3	-4	1	2	0	1	-2	-3	1	-2	-5	-1	0	0	-5	-3	-2	2		
Z Glx	0	0	1	3	-5	3	3	-1	2	-2	-3	0	-2	-5	0	0	-1	-6	-4	-2	2	3	
X ???	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

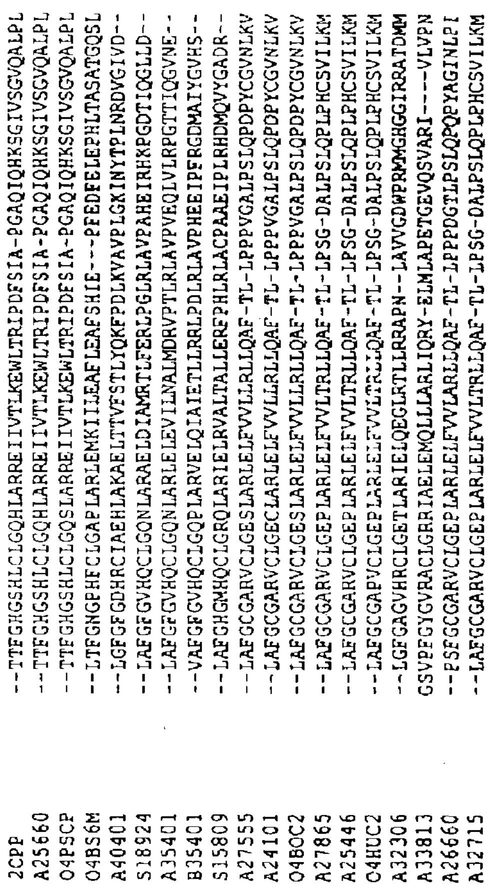
【図10】

FASTAによる結果ファイルを示す図（その2）



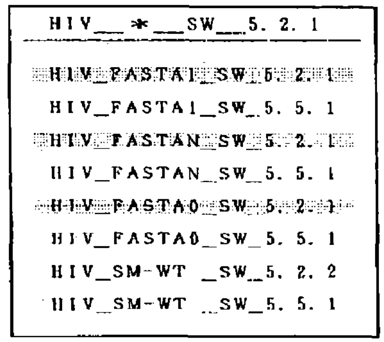
【図8】

配列データを示す図



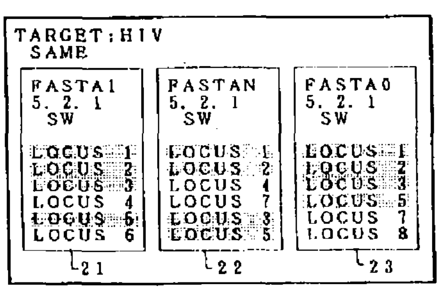
【図13】

結果ファイル名の画面表示を示す図



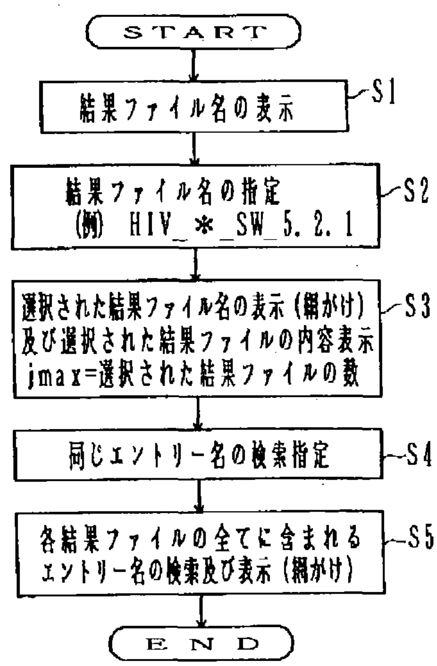
【図14】

共通する3配列名の画面表示を示す図



【図15】

共通する配列名の表示処理のフローチャート



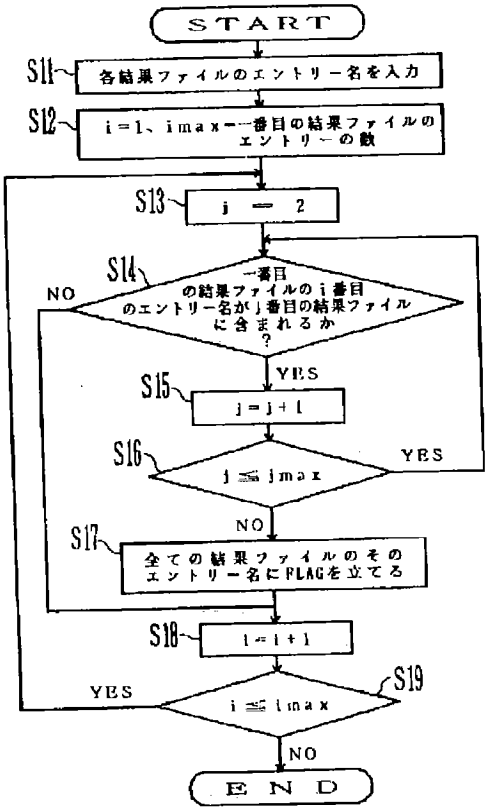
【図12】

FASTAによる結果ファイルを示す図（その4）

>O4B0C2	Steroid 21-monooxygenase cytochrome P450	111	76	78
>A27865	Steroid 21-monooxygenase cytochrome P450	111	71	80
>A25446	Steroid 21-monooxygenase cytochrome P450	111	71	80
>O4HUC2	Steroid 21-monooxygenase cytochrome P450	109	71	80
>A32106	Cytochrome P450 103 - Agrobacterium tumef	109	58	248
>A33813	Cytochrome P450 27 precursor, mitochondri	103	62	65
>A26660	Steroid 21-monooxygenase cytochrome P450	102	60	85
>A32715	Steroid 21-monooxygenase cytochrome P450	102	62	62
>JQ1143	Thromboxane-A synthase - Human (EC-number	97	59	68
>A39740	*Sterol 27-hydroxylase precursor - Human	90	55	57
>MNWVRA	Nonstructural polyprotein - Ross River vi	89	60	60
>A29587	Steroid 17alpha-monooxygenase cytochrome	83	55	104
>A40921	*Steroid 17alpha-monooxygenase - Human (E	83	55	104
>A40908	*Steroid 17alpha-monooxygenase (cytochrom	83	55	104
>A26366	Steroid 17alpha-monooxygenase cytochrome	83	55	104
>S12969	*Diacylglycerol kinase - Human (EC-number	80	41	42
>A26289	Steroid 17alpha-monooxygenase cytochrome	78	51	82
>A39607	*Mutation suppressor protein SRP3-1 - Yea	78	35	45
>S04346	Steroid 17alpha-monooxygenase cytochrome	78	51	89
>GNVVTR	Genome polyprotein - Tomato ringspot viru	77	45	46
>QQBE47	DNA-binding protein - Human herpesvirus 4	77	49	55
>TDHULX	Leukocyte antigen-related protein precurs	75	55	73
>B35342	Steroid 11beta-monooxygenase 2 cytochrome	75	50	81
>S09736	*Aldosterone synthase - Rat	75	50	81
>A32693	*Steroid receptor protein svp 1 - Fruit f	75	42	51
>A35342	Steroid 11beta-monooxygenase 1 cytochrome	75	50	81
>MNWVS	Nonstructural polyprotein - Sindbis virus	74	64	64
>VCLJG4	env polyprotein - Simian immunodeficiency	74	41	41
>GNNYBT	Genome polyprotein - Coxsackievirus B3	73	59	59
>GNNYB3	Genome polyprotein - Coxsackievirus B3	73	59	59
>S09156	Diacylglycerol kinase, lymphyocyte - Pig	73	35	36
>A35867	Cytochrome P450 71 - Avocado	73	43	49
>A41039	*RNA-directed RNA polymerase - Rabbit hem	73	41	51
>A27124	H+-transporting ATPase - Leishmania donov	73	41	52
>A32525	Steroid 21-monooxygenase cytochrome P450	72	60	84
>MNWVB2	Nonstructural polyprotein - Ockelbo virus	71	61	61
>DJBE28	DNA-directed DNA polymerase - Human herpe	71	41	41
>JX0050	Steroid 11beta-monooxygenase 3 cytochrome	69	41	41
>JX0071	Steroid 11beta-monooxygenase (clone 7-1)	69	41	41
>A38819	*Steroid 11beta-monooxygenase 2 (cytochro	69	41	41
>B26366	Steroid 17alpha-monooxygenase cytochrome	69	51	77
>JX0151	Steroid 11beta-monooxygenase 3 (cytochrom	69	41	41
>A28415	Steroid 11beta-monooxygenase cytochrome P	69	41	41
>S15805	*Cytochrome P450(11beta) - Bovine	69	41	41
>KIECG	GTP pyrophosphokinase - Escherichia coli	69	53	60
>A29943	Toll protein precursor - Fruit fly (Droso	69	43	48
>NCEC7	Exodeoxyribonuclease VII large chain - Es	68	57	66
>C32575	*C-ski protein FB27 - Chicken	68	53	53
>B41378	*Cytochrome c553i precursor - Paracoccus	68	68	79
>TVFVSK	Transforming protein (ski) - Avian erythr	68	53	53
>S18188	*Rat notch protein - Rat	68	43	43
>A40043	*Notch protein homolog TAN-1 precursor -	68	43	43
>A32575	*C-ski protein FB29 - Chicken	68	53	53
>A22363	Cytochrome P450, phenobarbital-inducible	68	46	51
>S16873	Cytochrome P450 2D4 - Rat	67	39	87
>A40457	*Replication protein A 70K chain - Human	67	67	76
>B37222	Cytochrome P450 1A2, hepatic - Dog (fragm	67	57	91
>S15435	*Collagen alpha 1(VIII) chain - Human	66	50	58
>A39129	*Catalase HPII - Escherichia coli (EC-num	66	55	58
>A34246	Collagen alpha 1(VIII) chain precursor -	66	50	62
>S05962	Radial spoke protein 3 - Chlamydomonas re	65	39	39
>S13178	*6-Methylsalicylate decarboxylase - Penic	65	41	46
>A40440	*Endothelin 1 and 2 receptor precursor, v	65	43	45
>A25076	Tubulin alpha-1 chain - Yeast (Saccharomy	65	39	40
>S15074	*Calpastatin - Rat	65	50	51
>A31270	*Radial spoke protein 3 - Chlamydomonas r	65	39	39

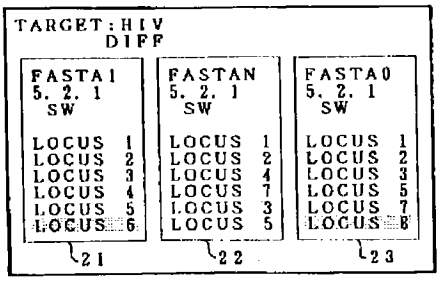
【図１６】

共通する配列名の抽出処理のフローチャート



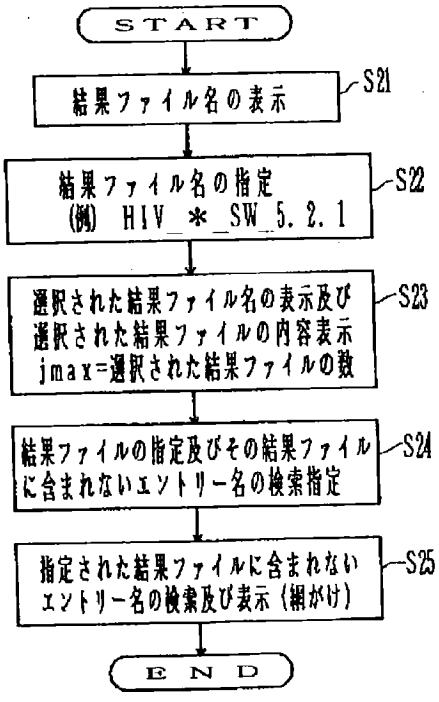
【図１７】

指定された結果ファイルに含まれない配列名の画面表示を示す図



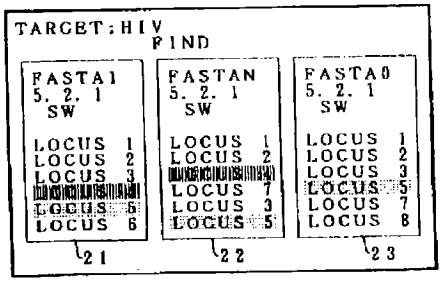
【図１８】

指定された結果ファイルに含まれない配列名の表示処理のフローチャート



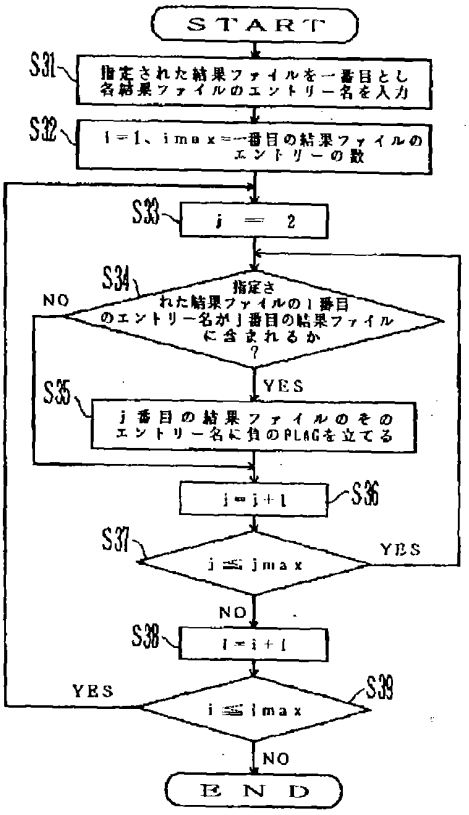
【図２３】

指定された配列名の画面表示を示す図



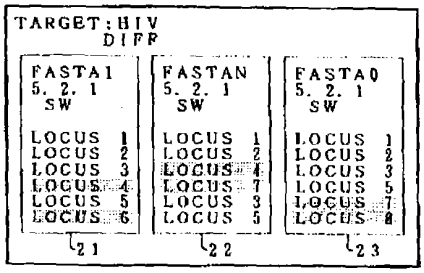
【図１９】

指定された結果ファイルに含まれない配列名の抽出処理のフローチャート



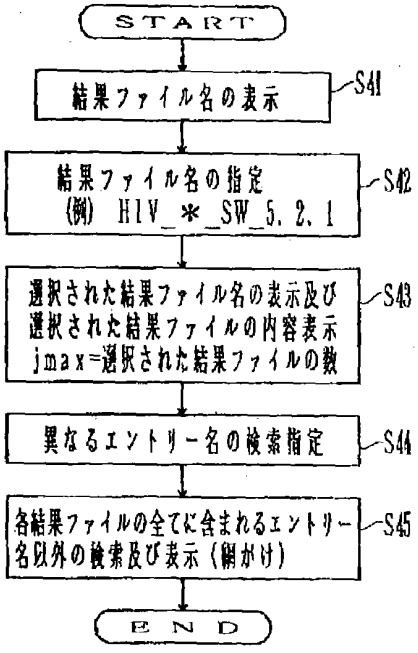
【図２０】

共通して含まれない配列名の画面表示を示す図



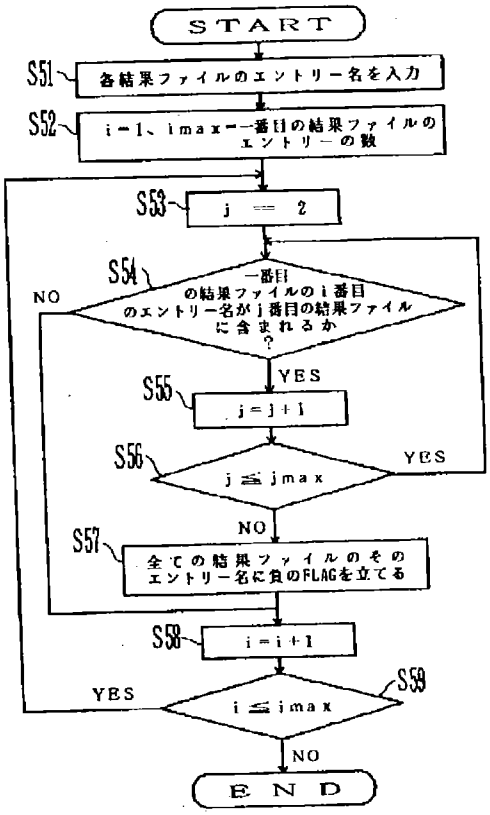
【図２１】

共通して含まれない配列名の表示処理のフローチャート



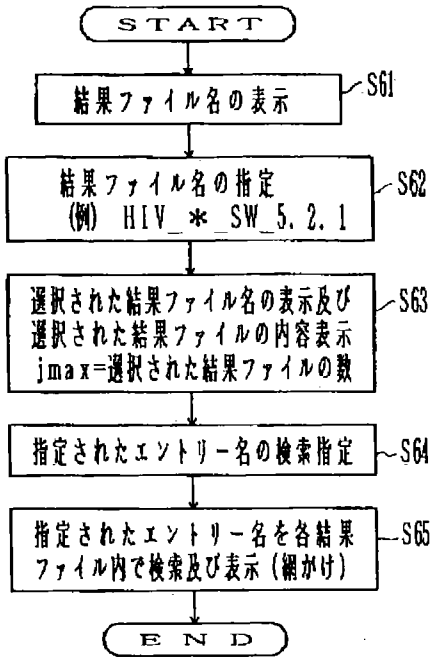
【図２２】

共通して含まれない配列名の抽出処理のフローチャート



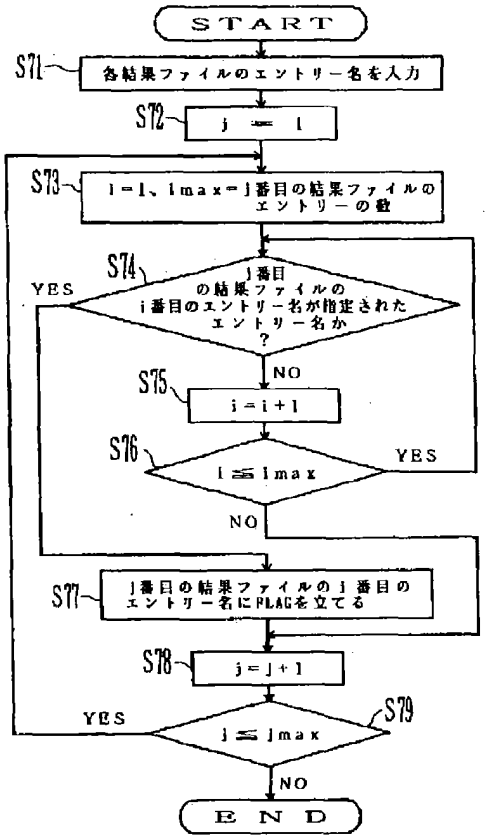
【図２４】

指定された配列名の表示処理のフローチャート



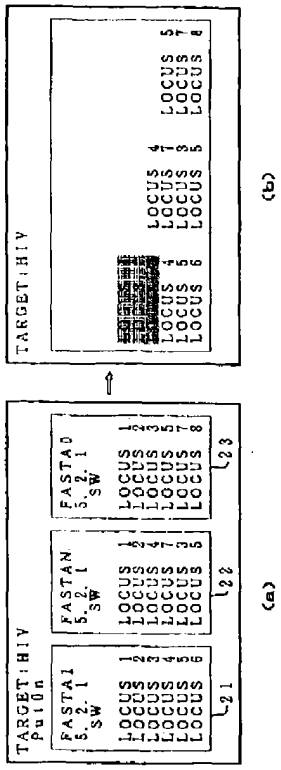
【図２５】

指定された配列名の抽出処理のフローチャート



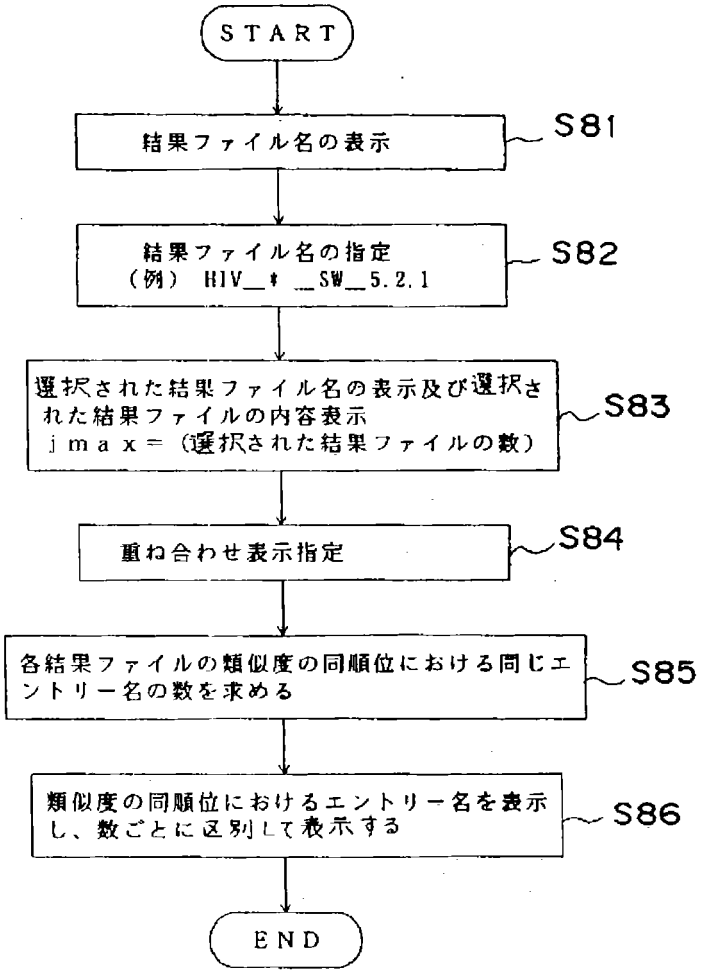
【図２６】

共通する同領位の配列名の画面表示を示す図



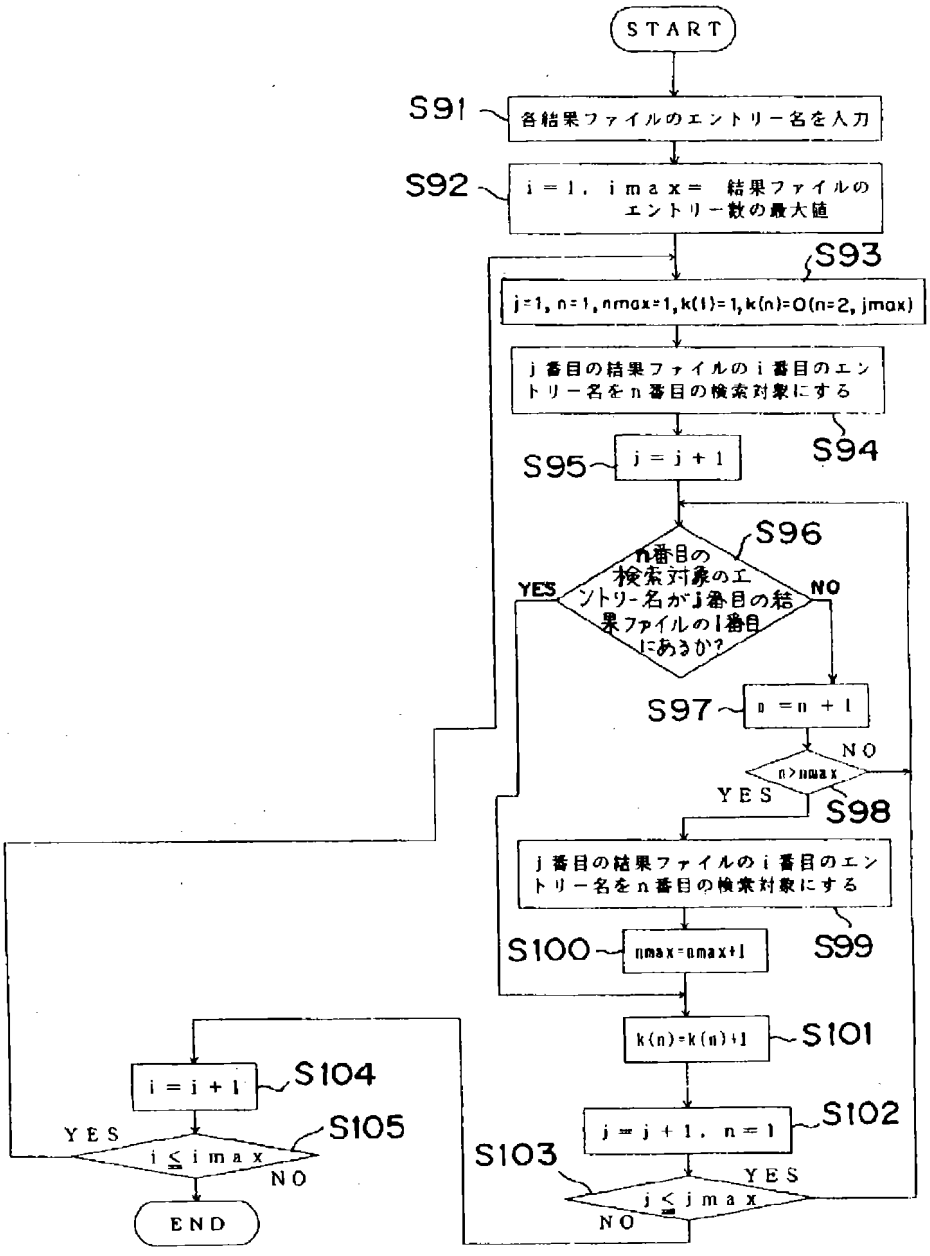
【図２７】

共通する同順位の配列名の表示処理のフローチャート



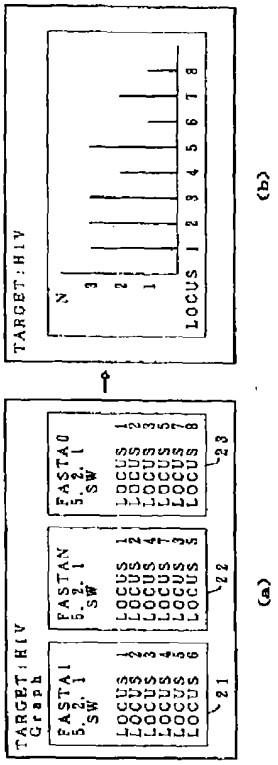
【図28】

共通する同順位の配列名の抽出処理のフローチャート



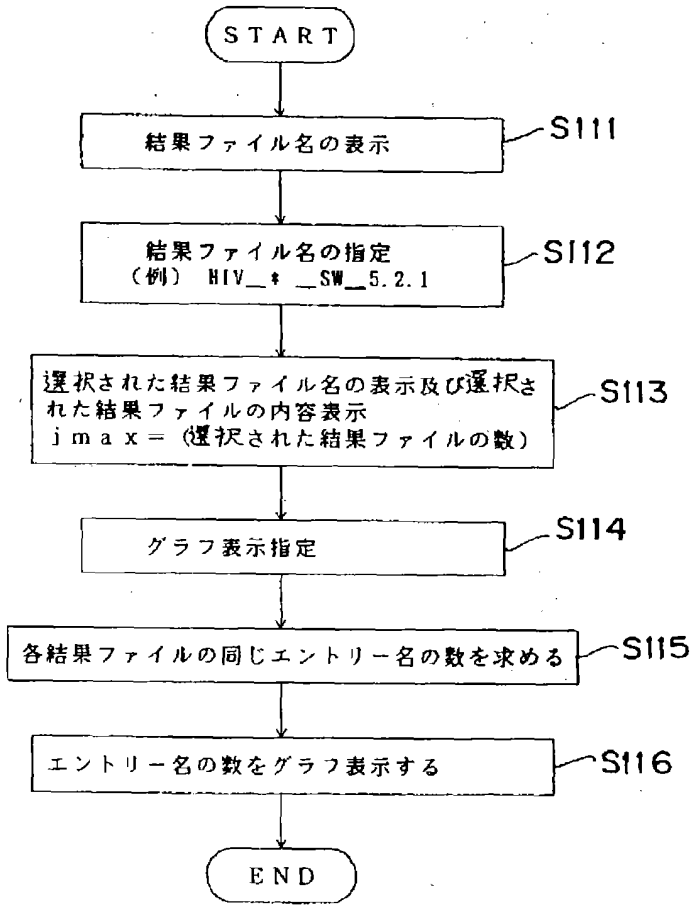
【図２９】

配列名の数のグラフ表示を示す図



【図３０】

配列名の数のグラフ表示処理のフローチャート



【図31】

配列名の数の計算処理のフローチャート